

ATTRACTIVE REGULAR STOCHASTIC CHAINS: PERFECT SIMULATION AND PHASE TRANSITION

SANDRO GALLO AND DANIEL YASUMASA TAKAHASHI

ABSTRACT. We prove that uniqueness of the stationary chain compatible with an attractive regular probability kernel is equivalent to the following two assertions for this chain: (1) it is a finitary coding of an i.i.d. process with discrete state space, (2) the concentration of measure holds at exponential rate. We show in particular that if a stationary chain is uniquely defined by a kernel which is continuous and attractive, then this chain can be sampled using a coupling-from-the-past algorithm. For the original Bramson-Kalikow model we further prove that there exists a unique compatible chain if and only if the chain is a finitary coding of a finite alphabet i.i.d. process. Finally, we obtain some partial results on conditions for phase transition for general chains of infinite order.

1. INTRODUCTION

In this work we consider chains of infinite order on finite alphabet, which are families of processes specified by kernels of transition probabilities that can depend on the whole past. This general class of processes includes the finite order Markov chains as special cases, but also includes stochastic models that exhibit phase transitions. An important question in the area is “what properties distinguish kernels exhibiting phase transition from kernel satisfying uniqueness?”. This work gives necessary and sufficient conditions for the existence of phase transition for an important class of chains of infinite order, namely for the attractive regular chains.

A probability kernel is called regular when it is continuous with respect to the past and has no null transition probabilities. The regularity of the kernel guarantees the existence of at least one chain compatible with the kernel. Attractiveness means that there exists some monotonicity property for the transition probabilities and it is analogous to the attractiveness of specification considered in the statistical mechanics (Preston, 1976).

It came with some surprise when Bramson & Kalikow (1993) showed an example of family of regular and attractive kernels with more than one compatible chain. Some interesting results are based on this Bramson-Kalikow (BK) example. For instance, Quas (1996) used the BK example to construct a C^1 expanding map of the circle which preserves Lebesgue measure such that the system is ergodic, but not weak-mixing. Also using the BK example, Stenflo (2001) showed a counterexample to a conjecture raised by Karlin (1953). Lacroix (2000) obtained some simplification to the Bramson & Kalikow (1993) proof of phase transition and Hulse (2006) showed a different example of regular and attractive kernel exhibiting phase transition using similar ideas of proofs as

2000 *Mathematics Subject Classification.* Primary 60G10; Secondary 60G99.

Key words and phrases. Chains of infinite order, perfect simulation, phase transition, finitary coding.

SG was supported by FAPESP grant 2009/09809-1. DYT was supported by FAPESP grant 2008/08171-0 and Pew Latin American Fellowship. This work is part of USP project “Mathematics, computation, language and the brain.”

in Lacroix (2000). To best of our knowledge, Berger *et al.* (2005) exhibited the only non-attractive example of non-unique chain of infinite order and is not based on BK example. Despite the importance of these works, none of them give general sufficient conditions for the existence of phase transition, even for the special (but important) case of attractive kernels.

In the present work, we prove that, for attractive regular kernels, uniqueness of the stationary chain compatible with a kernel is equivalent to the following two assertions: (1) the compatible chain is a finitary coding of a discrete-value i.i.d. process, (2) the concentration of measure holds at exponential rate. The equivalence (1) means that, for attractive regular chains, uniqueness implies that the compatible chain is a factor of a discrete-value i.i.d. process whose mapping depends almost surely on a finite number of coordinates with respect to the measure of the i.i.d. process. Equivalence (2) means that phase transition yields loss of “good” concentration of measure. We show that the regularity of the kernel is essential for the three equivalences to hold and, in general, cannot be relaxed. We also obtain some partial results for non-regular and non-attractive cases, which are of independent interest.

The main ingredient for the proof of the existence of a finitary coding is the proof that uniqueness of compatible chain for the regular and attractive kernels is equivalent to the existence of a Coupling-from-the-Past (CFTP) perfect simulation algorithm. This class of simulation algorithms was first introduced for the Markov chains by Propp & Wilson (1996) and then generalized to several other stochastic models. It is interesting that despite its simplicity, our algorithm can generate samples of continuous and attractive chains under regimes in which it was previously not known to be possible.

Finally, from a coding point of view, it is interesting to have a finitary coding from a finite alphabet i.i.d. process. We show that this is possible in the original Bramson & Kalikow (1993) example, if and only if there is a unique compatible chain, that is, choosing the parameters of the model in such a way that uniqueness holds. The proof of this fact is done by explicitly constructing a coding function that yields a finitary coding from an i.i.d. chain with finite alphabet to an i.i.d. chain with countable alphabet, if the later has finite entropy. This result and the proof technique has an interest of its own.

This article is organized as follows. In Section 2 we introduce the notation, definitions and the necessary backgrounds, with examples, to state the main results in Section 3. In Section 4 we discuss the results and some of their implications. In Section 5 we introduce the Attractive Sampler, which is used to prove the theorems of Section 3. Finally in Section 6 we prove all the results.

2. NOTATION, STANDARD DEFINITIONS AND EXAMPLES

Notation. For any set \mathcal{U} we denote the sets of bi-infinite, left-infinite and finite sequences of symbols of \mathcal{U} by $\mathcal{U}^{-\mathbb{Z}} = \mathcal{U}^{\{\dots, 2, 1, 0, -1, -2, \dots\}}$, $\mathcal{U}^{-\mathbb{N}} = \mathcal{U}^{\{-1, -2, \dots\}}$ and $\mathcal{U}^* = \bigcup_{j \geq 1} \mathcal{U}^{\{-1, \dots, -j\}}$, respectively. The elements of these sets will be denoted, respectively, $\mathbf{u} = \dots u_2 u_1 u_0 u_{-1} u_{-2} \dots$, $\underline{u} = u_{-1} u_{-2} \dots$ and $u_{-k}^{-1} = u_{-1} u_{-2} \dots u_{-k}$ for any $1 \leq k < +\infty$.

In the present paper \mathcal{U} is some Polish space and A is the finite ordered set $\{1, 2, \dots, s\}$ unless specified. A is called *alphabet*. We define a partial order on $A^{-\mathbb{N}}$ by saying that $\underline{a} \leq \underline{b}$ whenever $a_{-i} \leq b_{-i}$ for any $i \geq 1$. In $A^{-\mathbb{N}}$, the maximal element is \underline{s} , and the minimal element is $\underline{1}$.

Chains of infinite order. A *probability kernel*, or simply kernel, P on alphabet A is a function

$$\begin{aligned} P : A \times A^{-\mathbb{N}} &\rightarrow [0, 1] \\ (a, \underline{x}) &\mapsto P(a|\underline{x}) \end{aligned}$$

such that

$$\sum_{a \in A} P(a|\underline{x}) = 1, \quad \forall \underline{x} \in A^{-\mathbb{N}}.$$

We say that a stationary stochastic chain $\mathbf{X} = \{X_j\}_{j \in \mathbb{Z}}$ (of stationary law μ) on A is *compatible* with a kernel P if the later is a regular version of the conditional probabilities of the former, that is

$$\mu(X_0 = a | X_{-\infty}^{-1} = \underline{x}) = P(a|\underline{x})$$

for every $a \in A$ and μ -a.e. \underline{x} in $A^{-\mathbb{N}}$. When there is more than one stationary chain compatible with P , we say that there is *phase transition*, otherwise we say that the chain is *unique*. We follow the Harris nomenclature (Harris, 1955) and call these chains *chains of infinite order*. These processes were first introduced by (Onicescu & Mihoc, 1935) under the name *chaînes à liaisons complètes*. The existence of an invariant measure for these chains was first studied by Doeblin & Fortet (1937). It was rediscovered several times under different names (Harris, 1955; Keane, 1972; Kalikow, 1990). For a comprehensive historical account and recent developments we refer the reader to Fernández & Maillard (2005).

Non-nullness, continuity rate, oscillations and attractiveness. We say that a kernel P is *strongly non-null* if

$$\inf_{a \in A, \underline{x} \in A^{-\mathbb{N}}} P(a|\underline{x}) > 0.$$

The *continuity rate* (or *variation*) of order k of a kernel P is

$$\text{var}_k := \sup_{b \in A} \sup_{a_{-k}^{-1} \in A^k} \sup_{\underline{x}, \underline{y} \in A^{-\mathbb{N}}} |P(b|a_{-k}^{-1}\underline{x}) - P(b|a_{-k}^{-1}\underline{y})|.$$

We say that P is continuous if $\lim_{k \rightarrow \infty} \text{var}_k = 0$. A compactness argument shows that there always exists at least one stationary chain compatible with a continuous kernel (see for example Keane (1972)). If P is strongly non-null and continuous, we say that P is a *regular kernel*. Another characterization of kernels is the *oscillation rate*:

$$\text{osc}_n := \sum_{a \in A} \text{osc}_n(a) \quad \text{where} \quad \text{osc}_n(a) := \sup\{|P(a|\underline{x}) - P(a|\underline{y})| : \underline{x}, \underline{y} \in A^{-\mathbb{N}} \ x_{-i} = y_{-i} \ \forall i \neq n\}.$$

The sequences $\{\text{var}_k\}_{k \geq 0}$ and $\{\text{osc}_k\}_{k \geq 0}$ are related to the uniqueness of the compatible stationary chain as we will see in the examples below.

Finally, we say that a kernel P on A is *attractive* if for all $a \in A$ the value of $\sum_{j \geq a} P(j|\underline{x})$ is increasing on $\underline{x} \in A^{-\mathbb{N}}$.

Let us give two important examples taken from the literature, which we will repeatedly use in the sequel to illustrate our assertions.

Binary auto-regressive models. These models are extensively used in the statistical literature (McCullagh & Nelder, 1983), and are defined through the following parameters. A continuously differentiable increasing function $\psi : \mathbb{R} \rightarrow]0, 1[$, a summable sequence of non-negative real numbers $\{\xi_n\}_{n \geq 1}$, and a non-negative real parameter $\gamma \geq 0$. Consider the class of kernels P on alphabet $\{-1, +1\}$ such that

$$P(a|\underline{x}) := \psi \left(a \sum_{n \geq 1} \xi_n x_{-n} + a\gamma \right).$$

These kernels are attractive and regular. The continuity follows from the fact that ψ is continuous. The strongly non-nullness is due to the definition of ψ , and the fact that $\sum_{n \geq 1} \xi_n < +\infty$. It is also attractive because for any ordered pair $\underline{x} \leq \underline{y}$ we have $\sum_{n \geq 1} \xi_n x_{-n} \leq \sum_{n \geq 1} \xi_n y_{-n}$, and since ψ is an increasing function, it follows that $P(+1|\underline{x}) \leq P(+1|\underline{y})$.

If ψ is Lipschitz continuous, then, one directly obtains $\text{var}_k \leq C \sum_{n > k} \xi_n$ and $\text{osc}_n \leq C \xi_n$ for some positive constant C . The criteria of Johansson & Öberg (2003) and Fernández & Maillard (2005) imply, respectively, that $\xi_n = Cn^{-\alpha}$ with $\alpha > 3/2$ and $\sum_{n \geq 1} \xi_n < 1$ both imply uniqueness of the stationary chain compatible with P .

An important example of such models is when $\psi(r) = e^{-r}(e^{-r} + e^r)$. The resulting kernel is called *logit model* in the statistics literature, and *one-sided 1-dimensional long-range Ising model* in statistical physics literature. For instance, Hulse (2006) used this model to give an example of phase transition in chains of infinite order.

The Bramson & Kalikow (1993) example. Consider an increasing function $\phi : [-1, +1] \rightarrow]0, 1[$, an increasing sequence of positive integers $\{m_j\}_{j \geq 1}$ and a sequence $\{\lambda_j\}_{j \geq 1}$ such that $\lambda_j \geq 0$ and $\sum_{j \geq 1} \lambda_j = 1$. We call the BK example the class of kernels defined on $A = \{-1, +1\}$ as

$$P(+1|\underline{x}) = \sum_{j \geq 1} \lambda_j \phi \left(\frac{1}{m_j} \sum_{i=1}^{m_j} x_{-i} \right).$$

The kernels of this class are attractive and regular. Attractiveness and strongly non-nullness follow directly from the definition of ϕ and simple calculations yields $\text{var}_n \leq \sum_{\{j: m_j > n\}} \lambda_j$ showing that P is indeed continuous as well. When $\phi(s) = (1 - \epsilon)\mathbf{1}\{s > 0\} + \epsilon\mathbf{1}\{s < 0\}$ for some $\epsilon \in (0, 1)$ and $\lambda_k = cr^{-k}$ for some $r \in (2/3, 1)$ and $c = (1 - r)/r$, we call this model the *original BK* example, as it is precisely the model introduced in Bramson & Kalikow (1993). They proved that, taking the sequence $\{m_k\}_{k \geq 1}$ increasing sufficiently fast, the kernel P exhibits phase transition.

Maximum and minimum phases for attractive kernels. Define, for any $\underline{x} \in A^{-\mathbb{N}}$, the fixed past chain $\mathbf{X}^{\underline{x}}$ by

$$X_n^{\underline{x}} = \begin{cases} a_n & \text{if } n \leq 0 \\ a & \text{w.p. } P(a|X_{n-1}^{\underline{x}} \dots X_1^{\underline{x}} \underline{x}) \text{ otherwise.} \end{cases} \quad (1)$$

$\mathbf{X}^{\underline{x}}$ is a the non-stationary chain obtained by fixing the past \underline{a} from time $-\infty$ to 0, and “running P from this past”. For any \underline{x} , $j \geq 1$, and $n \in \mathbb{Z}$, let $\mathbf{X}^{\underline{x}, j}$ be the process defined by $X_n^{\underline{x}, j} := X_{n+j}^{\underline{x}}$. The attractiveness of P implies that for $\underline{x} = \underline{1}$ and \underline{s} , the sequence of processes $\{\mathbf{X}^{\underline{x}, j}\}_{j \geq 1}$ is stochastically non-decreasing and non-increasing, respectively (Hulse, 1991), and therefore, the weak limits

$$\mathbf{X}^{\min} := \lim_{j \rightarrow +\infty} \mathbf{X}^{\underline{1}, j} \text{ and } \mathbf{X}^{\max} := \lim_{j \rightarrow +\infty} \mathbf{X}^{\underline{s}, j} \quad (2)$$

exist and are stationary. If P is continuous, then \mathbf{X}^{\min} and \mathbf{X}^{\max} are compatible with P . We call \mathbf{X}^{\min} the *minimum phase* and \mathbf{X}^{\max} the *maximum phase*.

Finitary process. Let the shift operators $T_{\mathcal{U}}$ and T_A acting respectively on $\mathcal{U}^{\mathbb{Z}}$ and $A^{\mathbb{Z}}$ by shifting the sequences of one unit: $T_{\mathcal{U}}(\mathbf{u}) = \{u_{i+1}\}_{i \in \mathbb{Z}}$ and $T_A(\mathbf{a}) = \{a_{i+1}\}_{i \in \mathbb{Z}}$. A stationary process \mathbf{X} (with stationary law μ) on alphabet A is a *stationary coding* of a stationary process \mathbf{U} (with stationary law \mathbb{P}) on \mathcal{U} if there exists a measurable function $\Phi : \mathcal{U}^{\mathbb{Z}} \rightarrow A^{\mathbb{Z}}$ which is *translation equivariant* (that is $\Phi(T_{\mathcal{U}}(\mathbf{U})) = T_A \Phi(\mathbf{U})$) and such that $\mu = \mathbb{P} \circ \Phi^{-1}$. We follow the nomenclature given in Shields (1996) and call *B-processes* the processes that are stationary coding of an i.i.d. process. We have a *finitary coding* if there exists a *stopping time* $\theta : \mathcal{U}^{\mathbb{Z}} \rightarrow \mathbb{N} \cup \{\infty\}$, \mathbb{P} -a.s. finite, such that

$$[\Phi(\mathbf{U})]_0 = [\Phi(\mathbf{V})]_0 \text{ whenever } U_{-\theta(\mathbf{U})}^{+\theta(\mathbf{U})} = V_{-\theta(\mathbf{U})}^{+\theta(\mathbf{U})}. \quad (3)$$

This last assumption amounts to say that the event $\{\theta(\mathbf{U}) = k\}$ is $\mathcal{F}(U_{-k}^{+k})$ -measurable, or in other words, that the stopping time θ is checkable looking only at a finite number of U_i 's. We call *finitary processes* (FP) the processes that are finitary coding of an i.i.d. process. The notion of stationary coding comes from ergodic theory, and has a one-side analogue in the literature of stochastic processes, called the *Coupling from the past* algorithm (*CFTP algorithm* in the sequel). Such algorithms (which were first introduced in Propp & Wilson (1996) for Markov chains) aim to construct the function Φ using only the past values of an i.i.d. process \mathbf{U} , and the kernel P . If a CFTP algorithm is possible for a given kernel P , then the stationary measure μ which is constructed is a FP, because it is a particular finitary coding of \mathbf{U} , for which the event $\{\theta(\mathbf{U}) = k\}$ is $\mathcal{F}(U_{-k}^0)$ -measurable. In the literature, sometime a process is called finitary only if the set \mathcal{U} is finite. We do not assume this. In the special case of \mathcal{U} finite (or discrete), we say that the process is a *finitary coding of a finite (discrete) alphabet i.i.d. process*.

Exponential concentration of measure. Let $f : A^n \rightarrow \mathbb{R}$ be measurable. Define $\delta_j f = \sup\{|f(a_1^n) - f(b_1^n)| : a_i = b_i, \forall i \neq j\}$ and let δf be the vector with j -th coordinate given by $\delta_j f$. We say that the *concentration of measure holds at exponential rate* for a stationary process \mathbf{X} if, for all $n > 0$ and functions f such that $\|\delta f\|_{\ell_1(\mathbb{N})} \leq \gamma < \infty$, we have

$$\mathbb{P}(|f(X_1^n) - \mathbb{E}[f(X_1^n)]| > \epsilon) \leq \exp\left(-\frac{g(\epsilon, \gamma)}{\|\delta f\|_{\ell_2(\mathbb{N})}^2}\right)$$

with $g(\epsilon, \gamma) > 0$.

In particular, the above inequality implies that for all $k \in \mathbb{N}^*$ and $h : A^k \rightarrow [0, 1]$ we have

$$\mathbb{P}\left(\left|\frac{1}{n-k+1} \sum_{j=0}^{n-k} h(X_{j+1}^{j+k}) - \mathbb{E}[h(X_1^k)]\right| > \epsilon\right) \leq e^{-ng_k(\epsilon)} \quad (4)$$

where $g_k(\epsilon) > 0$.

3. MAIN RESULTS

Theorem 1. *Let P be an attractive regular kernel. The following are equivalent:*

- (1) *There exists an unique chain compatible with P .*
- (2) *\mathbf{X}^{\max} is a finitary coding of a discrete-value i.i.d. process.*
- (3) *The ergodic theorem holds at exponential rate for \mathbf{X}^{\max} .*

From Hulse (1991) we know that the maximum phase (*resp.* the minimum phase) is always B-process regardless of been equal or different from the minimal phase (*resp.* the maximum phase). Therefore, the fact to be a B-process does not distinguish the presence or not of phase transition. Theorem 1 shows that, for a regular attractive kernel P , the property to be a finitary coding of an i.i.d process distinguishes between existence or not of phase transition.

Another interesting consequence of the above theorem is that for an attractive and regular kernel exhibiting phase transition, the maximum phase has the ergodic theorem holding at subexponential rate.

We now show that Theorem 1 is optimal in the class of attractive chains, in the sense that if we relax either continuity or strong non-nullness, we can find examples of stationary chains that are FP and with ergodic theorem holding at exponential rate, although they are not uniquely determined by their conditional probabilities.

Relaxing the strong non-nullness assumption. The following example shows that in general we cannot relax the strong non-nullness condition. Before giving our example, we need some more definitions. For any $\underline{x} \in A^{-\mathbb{N}}$, let

$$\ell(\underline{x}) := \min\{j \geq 0 : x_{-j-1} = -1\}.$$

When we look backward in \underline{x} , $\ell(\underline{x})$ counts the number of $+1$ before finding the first -1 . We use the convention that $\ell(+1) = \infty$. Let $\{p_i\}_{i \geq 0}$ be a monotonically decreasing sequence of $[0, 1]$ -valued real numbers, and let $p_\infty = \lim_{i \rightarrow +\infty} p_i$. The kernel P is defined on $\{-1, +1\}$ by $P(-1|\underline{x}) = p_{\ell(\underline{x})}$ for any $\underline{x} \neq +1$ and $P(-1|+1) := p_\infty$. It is clear that this example is attractive. It is also continuous. To see this, observe that

$$\sup_{\underline{y}, \underline{z} \in A^{-\mathbb{N}}} |P(a|a_{-k}^{-1}\underline{y}) - P(a|a_{-k}^{-1}\underline{z})| = 0$$

for any $a_{-k}^{-1} \in A^k$, except for $(+1)^k$. Hence,

$$\sup_{a_{-k}^{-1} \in A^k} \sup_{\underline{y}, \underline{z} \in A^{-\mathbb{N}}} |P(a|a_{-k}^{-1}\underline{y}) - P(a|a_{-k}^{-1}\underline{z})| = \sup_{\underline{y}, \underline{z} \in A^{-\mathbb{N}}} |P(a|(+)^k \underline{y}) - P(a|(+)^k \underline{z})|,$$

and thus we obtain that

$$\sup_{a_{-k}^{-1} \in A^k} \sup_{\underline{y}, \underline{z} \in A^{-\mathbb{N}}} |P(a|a_{-k}^{-1}\underline{y}) - P(a|a_{-k}^{-1}\underline{z})| = p_k - p_\infty,$$

which goes to 0 by the definition of the sequence $\{p_k\}_{k \geq 0}$. If $p_\infty = 0$, the chain is not strongly non-null, and the degenerated chain with all symbols equal to $+1$ is trivially stationary and compatible with P . If we further assume $\sum_{k \geq 1} \prod_{i=0}^{k-1} (1 - p_i) < +\infty$, there exists another class of stationary chains compatible with the kernel P . It is the so-called class of renewal chains, obtained by concatenating i.i.d. blocks of the form $(+1, +1, \dots, +1, -1)$ of random length. These blocks have length $k+1$ with probability $\prod_{i=0}^{k-1} (1 - p_i)p_k$, and therefore, have finite expected length. The existence of several phases is due to the fact that this kernel is *not irreducible*. We refer to Cénac *et al.* (2010) for more details on this example. Therefore, this chain has a degenerate type of phase

transition. Nevertheless, the phase with probability one of having $+1$ is obviously a finitary coding of an i.i.d. process and concentration of measure at exponential rate. This shows that if the strong non-nullness assumption is removed, existence of finitary coding for the maximum phase does not imply uniqueness of the compatible chain.

Relaxing the continuity assumption. For a discontinuous attractive kernel P , the maximum and minimum phases are always distinct and not consistent with kernel P (see Hulse (1991)). Hence, strictly speaking, we don't have a phase transition where there is more than one chain compatible with P . In fact, this means that considering discontinuous attractive chains does not make much sense from the point of view of non-uniqueness. Nevertheless, we can still ask if the maximum and minimum phases of a discontinuous attractive kernel can be finitary coding of an i.i.d process. The example below shows that, in general, this could happen.

Let $\underline{a} \in \{-1, 1\}^{-\mathbb{N}}$ and let $-\underline{1}$ and $+\underline{1}$ be the pasts such that $+\underline{1}_j = +1$ and $-\underline{1}_j = -1$ for all j . Let $\{\epsilon_n\}_{n \in \mathbb{N}}$ be a non-increasing sequence of positive numbers such that $\lim_{n \rightarrow \infty} \epsilon_n = \epsilon$ with $0 < \epsilon < 1/2$. We define P by

$$\begin{aligned} P(1|\underline{a}^{-1}_{-n}(+\underline{1})) &= 1 - \epsilon_n = P(-1|-\underline{a}^{-1}_{-n}(-\underline{1})) \\ P(1|\underline{a}^{-1}_{-n}(-\underline{1})) &= \epsilon_n = P(-1|-\underline{a}^{-1}_{-n}(+\underline{1})), \end{aligned}$$

and put $P(1|\underline{x}) = 1/2$ for all the remaining pasts \underline{x} . Clearly P is strongly non-null, attractive, symmetric i.e. $P(1|\underline{z}) = P(-1|-\underline{z})$ and, non-continuous. Also, let

$$\begin{aligned} P_+(1|\underline{a}) &= \lim_{n \rightarrow \infty} P(1|\underline{a}^{-1}_{-n}(+\underline{1})) = 1 - \epsilon \\ P_-(-1|\underline{a}) &= \lim_{n \rightarrow \infty} P(-1|-\underline{a}^{-1}_{-n}(-\underline{1})) = 1 - \epsilon. \end{aligned}$$

By Lemma 2.3 in Hulse (1991), the maximum phase $\mathbf{X}^{(+1)}$ is consistent with P_+ and therefore it is the i.i.d. process with probability $1 - \epsilon$ for 1 . Analogously, the minus phase $\mathbf{X}^{(-1)}$ is the i.i.d. process with probability $1 - \epsilon$ for -1 .

Theorem 1 is a direct consequence of Theorems 2, 3 and 4 below.

Theorem 2. *Let P be an attractive continuous kernel. If there exists a unique stationary chain compatible with P then this chain can be perfectly sampled by a CFTP algorithm using a discrete-value i.i.d. process.*

Observe that for this theorem we do not require strong non-nullness of the kernel. As an immediate application of Theorem 2, our *Attractive Sampler* given in Section 5 perfectly simulates binary autoregressive and BK processes introduced in Section 2 in their uniqueness regime. Notice that for any $\eta > 0$, we can exhibit, for the binary autoregressive and BK processes, kernels having continuity rate $\text{var}_k = O(1/k^\eta)$ for which the unique stationary chain compatible can be perfectly simulated. In particular, in the BK example of Friedli (2010), var_k can be taken so that it converges arbitrarily slowly to 0. As a comparison, in the work of Comets *et al.* (2002) the condition $\sum_{k \geq 0} \prod_{i=0}^{k-1} (1 - \text{var}_k) = +\infty$ must be satisfied to guarantee the existence of a CFTP algorithm. This condition does not hold if $\text{var}_k = O(1/k^\eta)$ with sufficiently small η . In other words, in the class of regular attractive chains, our perfect simulation algorithm (Attractive Sampler) is optimal. This is particularly clear in the linear case of Example 1 which is also considered in Comets *et al.*

(2002). In this case, when $\sum_{n \geq 1} \xi_n + \gamma < 1$, the criterium of Fernández & Maillard (2005) implies uniqueness, and therefore, our Attractive Sampler works whereas their algorithm is not guaranteed to work.

The following theorem is closely related to Theorem 3 in Marton & Shields (1994).

Theorem 3. *Let \mathbf{X} be a FP with stopping time θ . Then the concentration of measure holds at exponential rate for \mathbf{X} .*

The above theorem holds for any process (we assume neither regularity nor attractiveness) that is FP with some stopping time θ and it is of independent interest. It is also interesting to note that the proof technique of the above theorem can be applied to obtain a sharper result for the particular case of process that has an CFTP algorithm as is the case for process considered for example in Comets *et al.* (2002), Gallo & Garcia (2011), and De Santis & Piccioni (2010). The result below gives an explicit upper bound for processes having a CFTP algorithm with finite expected stopping time.

Proposition 1. *Let \mathbf{X} be a process that can be simulated by a CFTP algorithm with a stopping time θ . If $\mathbb{E}[\theta] < \infty$, then for all $\epsilon > 0$ and all functions $f : A^n \rightarrow \mathbb{R}$ we have*

$$\mathbb{P}(|f(X_1^n) - \mathbb{E}[f(X_1^n)]| > \epsilon) \leq 2 \exp \left(- \frac{2\epsilon^2}{(1 + \mathbb{E}[\theta])^2 \|\delta f\|_{\ell_2(\mathbb{N})}^2} \right). \quad (5)$$

Now we have the last ingredient for the proof of Theorem 1.

Theorem 4. *Let P be a regular kernel and \mathbf{X} a process compatible with P that satisfies the ergodic theorem at exponential rate. Then \mathbf{X} is the unique stationary process compatible with P .*

Note that, for this result, we do not assume that the kernel is attractive and therefore, Theorem 4 constitutes an interesting characterization of uniqueness for chains of infinite order. We cannot, in general, relax the strong non-nullness condition in this theorem, because by the example given soon after Theorem 1 where the kernel has only one null transition probability, there exists a chain that satisfies the ergodic theorem at exponential rate but it is not the unique chain compatible with the kernel.

Now, Theorem 1 follows from the chain of implications shown in Figure 1 that holds when the kernel is attractive and regular.

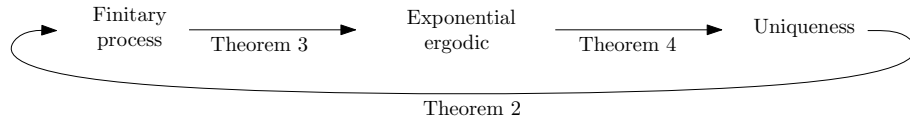


FIGURE 1. Diagram showing the chain of implications stated by Theorem 1

From the coding point of view, it is natural to ask if Theorem 1 can be strengthened to a finitary coding form a finite alphabet i.i.d. process. This is indeed the case for the original BK example.

Theorem 5. *Let P be the original BK example. Then there exists a unique chain compatible with P if and only if the compatible chain is a finitary coding of a finite alphabet i.i.d. process.*

4. DISCUSSION

In this work, we were motivated by the following general question: “what is the relationship between the existence of a CFTP perfect simulation algorithm and the occurrence of phase transition?” A similar question was studied in Steif & van den Berg (1999), where they proved that if a random field is obtained as invariant measures of monotonic, exponentially ergodic probabilistic cellular automata then the random field is a finitary coding of a finite alphabet i.i.d. random field. They proved the existence of a CFTP algorithm for this class of processes and then used this result to show that there exists a finitary coding from a finite alphabet i.i.d. process to the plus phase of a ferromagnetic Ising model below the critical temperature. They also proved that for the plus phase of an Ising model above the critical temperature, there is no finitary coding from a finite alphabet i.i.d. process.

It is natural to ask how much of this program can be carried out and improved for other stochastic processes. In this article, we consider the regular chains of infinite order, which are natural generalization of finite order Markov chains. This class of stochastic models has some similarities with Gibbs measures, although the theory developed for Gibbs measure cannot be applied directly to the chains of infinite order as there exist regular chains of infinite order that are not Gibbs (Fernández *et al.*, 2011). Also, it is fair to say that much less is known about chains of infinite order, when compared to the Gibbs measure. For instance, it is not known what is the behavior of the loss of memory for unique chains of infinite order in general even for specific model as the BK example, and therefore conditions like exponential ergodicity cannot be assumed a priori.

The main result of this article is Theorem 1, which gives a satisfactory answer to our general question, stating that, for attractive regular kernels, uniqueness of the compatible chain is equivalent to the existence of a CFTP algorithm using a discrete-value i.i.d. process which is itself equivalent to saying that the measure “enjoys good concentration”. For the proof of this theorem we have three ingredients. The first is the equivalence between uniqueness of compatible chain for attractive regular kernels and the existence of a CFTP perfect simulation algorithm for the compatible chain. As a corollary for this result, we obtain a perfect simulation algorithm for a large class of models previously not known to be possible. The second ingredient is the fact that if a process is a finitary coding of an i.i.d. process, then it has the concentration of measure at exponential rate. This result give us an explicit upper bound for the concentration of measure property based on the tail probability of the stopping time of the finitary coding. We note that the blowing up property (Marton & Shields, 1994) could be used instead of our concentration of measure result (Theorem 3), but we think that our result give us more explicit information about the process. Finally, the third ingredient is the proof that if a process compatible with a regular kernel has the ergodic theorem at exponential rate, then it is the unique compatible chain.

Petit (1982) proved that discrete-value i.i.d. processes with infinite entropy are finitary isomorphic. This implies also that given a discrete-value i.i.d. process with finite or infinite entropy, there exists a finitary coding from a unique discrete-value i.i.d. process with infinite entropy to the given i.i.d. process. Therefore, the finitary coding considered in this article can be made source universal, *i.e.*, the discrete-value i.i.d. process used to finitarily code does not depend on the particular attractive regular chains. Because the entropy of an attractive regular chain is finite but can be as large as possible depending on the choice of the model, there is no hope to obtain a

finitary coding using a universal finite alphabet i.i.d. process. Therefore, when considering source universal coding, our result is optimal with respect to the cardinality of the alphabet of the coding process.

For non-attractive and non-regular cases, we also obtain some results. Theorem 2 shows that for attractive continuous, not necessarily non-null kernels, uniqueness implies existence of a CFTP perfect simulation algorithm for the unique chain. Indeed, a routine argument shows that feasibility of our CFTP algorithm (Attractive Sampler) implies uniqueness. We note that this equivalence does not necessary hold for all CFTP algorithm as we exhibit an example of attractive continuous kernel with a single null transition probability that exhibit more than one compatible measure and for which the maximum phase is a finitary coding of an i.i.d. process. The important observation is that the coding is not realized using the Attractive Sampler for this example.

Theorem 3 shows that if a process, not necessarily attractive nor regular, can be sampled by a CFTP algorithm, then the concentration of measure holds at exponential rate. Therefore, combining Theorems 2 and 3, we have that if a process compatible with an attractive continuous kernel cannot have the ergodic theorem at exponential rate, then this kernel exhibit phase transition. We do not pursue further this observation in this work, but it seems interesting to investigate if one can prove the existence of phase transition studying rate of convergence of the ergodic theorem.

Finally, we show that the chain compatible with the original BK example is a finitary coding of a finite alphabet i.i.d. process if and only if it is unique. For the BK example, it is not clear for us what is the quantity equivalent to the statistical mechanics notion of temperature, and therefore we cannot make a one-to-one comparison with Theorem 1.1 in Steif & van den Berg (1999). Nevertheless, we observe that Steif & van den Berg (1999) was able to show, for the ferromagnetic Ising model, that the finitary coding from a finite alphabet was equivalent to uniqueness only up to the critical temperature, whereas Theorem 5 has no restriction.

5. THE ATTRACTIVE SAMPLER

Assume that we are given an attractive continuous kernel P for which there exists a unique compatible stationary chain. The alphabet is $A = \{1, \dots, s\}$. As stated by Theorem 2, there exists a CFTP algorithm that samples from its stationary law. Here we construct one such CFTP algorithm, we will call the *Attractive Sampler*. First, let us introduce the kernel \tilde{P} on $A \times A$ defined as

$$\tilde{P}((a' \geq a, b' \geq b)|(x, y)) := \sum_{i=a}^s P(i|\underline{x}) \wedge \sum_{i=b}^s P(i|\underline{y}), \quad (6)$$

for any pair of pasts \underline{x} and \underline{y} and any pair of symbols a and b in A . This kernel defines a coupling between the kernel $\{P(a|\underline{x})\}_{a \in A}$ and $\{P(a|\underline{y})\}_{a \in A}$. To see this, first observe that

$$\sum_{a=1}^s \tilde{P}((a, s)|(x, y)) = \tilde{P}((a' \geq 1, b' \geq s)|(x, y)) = P(s|\underline{y}),$$

secondly, observe that

$$\begin{aligned} \sum_{a=1}^s \tilde{P}((a, s-1)|(x, y)) &= \tilde{P}((a' \geq 1, b' \geq s-1)|(x, y)) - \tilde{P}((a' \geq 1, b' \geq s)|(x, y)) \\ &= [P(s-1|\underline{y}) + P(s|\underline{y})] - P(s|\underline{y}) \\ &= P(s-1|\underline{y}). \end{aligned}$$

Now, continuing the recursion, we show that $\sum_{a=1}^s \tilde{P}((a, b)|(x, y)) = P(b|y)$ for any $b \in A$. The same holds for the sum over b , that is $\sum_{b=1}^s \tilde{P}((a, b)|(x, y)) = P(a|x)$ for any $a \in A$. This means that \tilde{P} defines a coupling between the chains with respective fixed pasts.

A straightforward but tedious computation shows that \tilde{P} is indeed continuous and we can use the result of Kalikow (1990) stating that continuous kernels can be written as a countable mixture of Markovian kernel of increasing order. Formally, for \tilde{P} , there exists a sequence of non-negative numbers $\{\lambda_k\}_{k \geq 0}$ with $\sum_{k \geq 0} \lambda_k = 1$ and a sequence of Markov kernels $\{P^{[k]}\}_{k \geq 0}$, where $P^{[k]}$ is a k -steps Markov kernel, such that

$$\tilde{P}((a, b)|(x, y)) = \sum_{k \geq 0} \lambda_k P^{[k]}((a, b)|(x_{-k}^{-1}, y_{-k}^{-1})).$$

We will now use this representation of \tilde{P} to define our algorithm. For this, for any integer $k \geq 0$, let

$$\mathcal{R}^{[k]} = \{P^{[k]}((a' \geq a, b' \geq b)|(x_{-k}^{-1}, y_{-k}^{-1})) \in [0, 1] : (a, b) \in A^2, (x_{-k}^{-1}, y_{-k}^{-1}) \in A^{2k}\}.$$

Denote the elements of $\mathcal{R}^{[k]}$ by $r_{k1} < r_{k2} < \dots < r_{k|\mathcal{R}^{[k]}|} = 1$. Then, we define $q_{k1} = r_{k1}$ and, for $1 < j \leq |\mathcal{R}^{[k]}|$, we put $q_{kj} = r_{kj} - r_{k(j-1)}$.

Now, let $\mathbf{L} = \{L_j\}_{j \in \mathbb{Z}}$ be an i.i.d. process with values on \mathbb{N} such that for all $k \in \mathbb{N}$ and $j \in \{1, \dots, |\mathcal{R}^{[k]}|\}$

$$\mathbb{P}\left(L_0 = \sum_{m=-1}^{k-1} |\mathcal{R}^{[m]}| + j\right) = \lambda_k q_{kj},$$

where we define $\mathcal{R}^{[-1]} = \emptyset$.

We will use the i.i.d. process \mathbf{L} to define our algorithm. For this, we define the *update* function $F : A^{-2\mathbb{N}} \times \mathbb{N} \rightarrow A^2$ by

$$F((x, y), L_0) = (a, b), \tag{7}$$

if $L_0 = \sum_{m=-1}^{k-1} |\mathcal{R}^{[m]}| + j$ and if, for all $(c, d) \in A^2$,

$$P^{[k]}((a' \geq c, b' \geq d)|(x_{-k}^{-1}, y_{-k}^{-1})) \notin \left[r_{kj}, P^{[k]}((a' \geq a, b' \geq b)|(x_{-k}^{-1}, y_{-k}^{-1})) \right].$$

Let the concatenation of pairs of symbols be understood coordinatewise, *i.e.*, $(a, b)(c, d) = (ac, bd)$ whenever (ac, bd) is well defined. Using this notation, we define, for any $-\infty < k \leq l \leq +\infty$, the successive iterations of F as

$$F_{[k, l]}((x, y), L_k^l) = F(F_{[k, l-1]}((x, y), L_k^{l-1}) \dots F_{[k, k]}((x, y), L_k)(x, y), L_l), \tag{8}$$

where $F_{[k, k]}(x, U_k) := F(x, U_k)$. Notice that for any (x, y) and (a, b) ,

$$\mathbb{P}(F((x, y), L_0) = (a, b)) = \tilde{P}((a, b)|(x, y)),$$

and therefore we obtain a coupling $\{(X_i^x, X_i^y)\}_{i \geq 0}$ because

$$F_{[0, i]}((x, y), L_0^i) \stackrel{d}{=} (X_i^x, X_i^y).$$

Furthermore, we define for any $i \in \mathbb{Z}$, the random variable

$$\theta[i] := \min\{j \geq 0 : F_{[i-j, i]}((\underline{a}, \underline{s}), L_{i-j}^i) \in \{(1, 1), \dots, (s, s)\} \text{ for all } \underline{a} \in A^{-\mathbb{N}}\}, \tag{9}$$

and an important observation is that in the particular case of attractive chains,

$$\theta[0] = \min\{j \geq 0 : F_{[-j, 0]}((\underline{1}, \underline{s}), L_{-j}^0) \in \{(1, 1), \dots, (s, s)\}\}. \tag{10}$$

Finally, define the *reconstruction function* of time $i \in \mathbb{Z}$ by

$$[\Phi(\mathbf{U})]_i = F_{[-\theta[i], i]}((\underline{1}, \underline{s}), L_{-\theta[i]}^i). \quad (11)$$

It can be shown in a standard way (see for example Propp & Wilson (1996) for the Markovian case, or Comets *et al.* (2002) for chains of infinite order) that if $\theta[0]$ is \mathbb{P} -a.s. finite (that is, the Attractive Sampler is feasible), the sample $[\Phi(\mathbf{U})]_0$ is constructed according to the unique stationary measure compatible with P . In fact, the compatibility and the stationarity follow from the construction. The uniqueness follows from the loss of memory the chain inherits because of the existence of almost surely finite “regeneration times” $\theta[i]$ for any $i \in \mathbb{Z}$.

Thus, this is a finitary coding which is particular in that $\{\theta[0] = k\}$ is $\mathcal{F}(U_{-k}^0)$ -measurable. In Section 6.1 (proof of Theorem 2), we will prove that for attractive continuous P , the Attractive Sampler is feasible if and only if P has a unique compatible chain.

Algorithm 1 Attractive Sampler

```

1: Input:  $F$ ; Output:  $\theta[0]$ ,  $[\Phi(\mathbf{L})]_0$ 
2: Sample  $L_0$  with distribution  $\mathbb{P}$ 
3:  $i \leftarrow 0$ ,  $\theta[0] \leftarrow 0$ ,  $[\Phi(\mathbf{L})]_0 \leftarrow \star$ 
4: while  $F_{[-i, 0]}((\underline{1}, \underline{s}), L_{-i}^0) \notin \{(1, 1), \dots, (s, s)\}$  do
5:    $i \leftarrow i + 1$ 
6:   Sample  $L_{-i}$  with distribution  $\mathbb{P}$ 
7: end while
8:  $\theta[0] \leftarrow i$ 
9:  $[\Phi(\mathbf{L})]_0 \leftarrow F_{[-i, 0]}((\underline{1}, \underline{s}), L_{-i}^0)$ 
10: return  $\theta[0]$ ,  $[\Phi(\mathbf{L})]_0$ .
```

6. PROOF OF THE RESULTS

From Section 3 it is clear that Theorem 1 follows directly from Theorems 2, 3, and 4.

6.1. Proof of Theorem 2. First we need the following lemma proved for the regular and attractive kernel by (Hulse, 1991). Here, we drop the unnecessary non-nullness condition.

Lemma 1. *Let P be attractive and continuous and consider the function F defined in Section 5. If there exist a unique chain compatible with P , then for all $i \in A = \{2, \dots, s\}$,*

$$\lim_{n \rightarrow \infty} \mathbb{P}(F_{[0, n]}((\underline{1}, \underline{s}), L_0^n) \geq (i, 1)) - \mathbb{P}(F_{[0, n]}((\underline{1}, \underline{s}), L_0^n) \geq (1, i)) = 0. \quad (12)$$

Observation 1. *Remember that $F_{[0, n]}((\underline{1}, \underline{s}), L_0^n) \stackrel{d}{=} (X_n^1, X_n^s)$ and therefore, we have that for any $i \in A$, $\mathbb{P}(F_{[0, n]}((\underline{1}, \underline{s}), L_0^n) \geq (i, 1))$ gives the law of X_n^1 .*

Proof. For $a, b, c, d \in A$, we denote $(a, b) \geq (c, d)$ if $a \geq c$ and $b \geq d$.

Because P is attractive, for all $l_1, \dots, l_k, k \in \mathbb{N}$ and $a_1, \dots, a_k \in A$

$$\mathbb{P}(F_{[0, n+l_1]}((\underline{1}, \underline{s}), L_0^{n+l_1}) \geq (1, a_1), \dots, F_{[0, n+l_k]}((\underline{1}, \underline{s}), L_0^{n+l_k}) \geq (1, a_k))$$

and

$$\mathbb{P}(F_{[0, n+l_1]}((\underline{1}, \underline{s}), L_0^{n+l_1}) \geq (a_1, 1), \dots, F_{[0, n+l_k]}((\underline{1}, \underline{s}), L_0^{n+l_k}) \geq (a_k, 1)) \quad (13)$$

are respectively non-increasing and non-decreasing in $n \in \mathbb{N}$. Therefore, both sequences are convergent in $n \in \mathbb{N}$ and their limits when n diverges define stationary chains. By constructions, for all $b \in A$ and $\underline{a} \in A^{-\mathbb{N}}$ the respective chains are compatible with the kernels P^{\max} given by $P^{\max}(b|\underline{a}) = \lim_{n \rightarrow \infty} P(b|a_{-n}^{-1}\underline{s})$ and $P^{\min}(b|\underline{a}) = \lim_{n \rightarrow \infty} P(b|a_{-n}^{-1}\underline{1})$.

If the kernel P is continuous, $P^{\max} = P^{\min} = P$, and therefore the both chains are compatible with P . This implies that if there exists only one chain compatible with P , for $i = 2, \dots, s$, we have

$$\lim_{n \rightarrow \infty} \mathbb{P}(F_{[0,n]}((\underline{1}, \underline{s}), L_0^n) \geq (i, 1)) - \mathbb{P}(F_{[0,n]}((\underline{1}, \underline{s}), L_0^n) \geq (1, i)) = 0,$$

as we wanted. □

We now prove that, for attractive regular chains, convergence (12) implies that the unique stationary chain compatible with P can be sampled by a CFTP algorithm. We have to construct two functions Φ and θ that satisfy the conditions given in Section 2 for the definition of CFTP and FP.

We consider the respective functions used for the the Attractive Sampler defined in Section 5, and it only remains to prove that $\theta[0]$ is \mathbb{P} -a.s. finite. We have

$$\mathbb{P}(\theta[0] > n) = \mathbb{P}(\cap_{j=-n}^0 \{F_{[j,0]}((\underline{1}, \underline{s}), L_j^0) \notin \{(1, 1), \dots, (s, s)\}\}),$$

which yields, using first the attractiveness and then the translation invariance of \mathbf{L}

$$\begin{aligned} \mathbb{P}(\theta[0] > n) &= \mathbb{P}(F_{[-n,0]}((\underline{1}, \underline{s}), L_{-n}^0) \notin \{(1, 1), \dots, (s, s)\}) \\ &= \mathbb{P}(F_{[0,n]}((\underline{1}, \underline{s}), L_0^n) \notin \{(1, 1), \dots, (s, s)\}). \end{aligned}$$

Thus, we want to prove that

$$\lim_{n \rightarrow +\infty} \mathbb{P}(F_{[0,n]}((\underline{1}, \underline{s}), L_0^n) \notin \{(1, 1), \dots, (s, s)\}) = 0.$$

Due to the attractiveness, our coupling guarantees that for $a \in A$

$$\mathbb{P}(F_{[0,n]}((\underline{1}, \underline{s}), L_0^n) \geq (a, a)) = \mathbb{P}(F_{[0,n]}((\underline{1}, \underline{s}), L_0^n) \geq (a, 1)).$$

Taking the limit and using (12), we have for any $a \in A$

$$\begin{aligned} \lim_{n \rightarrow +\infty} \mathbb{P}(F_{[0,n]}((\underline{1}, \underline{s}), L_0^n) \geq (a, a)) &= \lim_{n \rightarrow +\infty} \mathbb{P}(F_{[0,n]}((\underline{1}, \underline{s}), L_0^n) \geq (a, 1)) \\ &= \lim_{n \rightarrow +\infty} \mathbb{P}(F_{[0,n]}((\underline{1}, \underline{s}), L_0^n) \geq (1, a)). \end{aligned}$$

Now, for any $a \in A$, let $\Gamma(a) = \{(i, j) \in A^2 : i < a \text{ and } j \geq a\}$. From the above equation, we have that

$$\lim_{n \rightarrow +\infty} \mathbb{P}(F_{[0,n]}((\underline{1}, \underline{s}), L_0^n) \in \Gamma(a)) = 0,$$

and this implies

$$\lim_{n \rightarrow +\infty} \mathbb{P}(F_{[0,n]}((\underline{1}, \underline{s}), L_0^n) \notin \{(1, 1), \dots, (s, s)\}) = 0,$$

which concludes the proof.

6.2. Proof of Theorem 3. Let \mathbf{X} be any process with alphabet A . For all $\underline{x} \in A^{-\mathbb{N}}$, let $\mathbf{X}^{\underline{x}}$ be the process with fixed past \underline{x} . Let $\underline{y}, \underline{z} \in A^{-\mathbb{N}}$ and $(\mathbf{X}^{\underline{y}}, \mathbf{X}^{\underline{z}})$ be a coupling between the process $\mathbf{X}^{\underline{y}}$ and $\mathbf{X}^{\underline{z}}$. The following lemma, which we state without proof, is a direct consequence of Theorem 1 of Chazottes *et al.* (2007).

Lemma 2 (Chazottes *et al.* (2007)). *Let $(\mathbf{X}^{\underline{y}}, \mathbf{X}^{\underline{z}})$ be couplings for each pairs $\underline{y}, \underline{z} \in A^{-\mathbb{N}}$. If $\sup_{\underline{y}, \underline{z}} \sum_{j=1}^{\infty} \mathbb{P}(X_j^{\underline{y}} \neq X_j^{\underline{z}}) \leq \Delta < \infty$, then for all integer $n \geq 1$, functions $f : A^n \rightarrow \mathbb{R}$, and $\epsilon > 0$ we have*

$$\mathbb{P}(|f(X_1^n) - \mathbb{E}[f(X_1^n)]| > \epsilon) \leq 2 \exp \left(- \frac{2\epsilon^2}{(1 + \Delta)^2 \|\delta f\|_{\ell_2(\mathbb{N})}^2} \right).$$

Assume that the process \mathbf{X} is a finitary coding of a sequence \mathbf{U} , where $U_i \in \mathcal{U}$ for any $i \in \mathbb{Z}$. Let θ and Φ be the quantities involved in the finitary coding as defined generically in Section 2 (and not necessarily as in Section 5). Let $\epsilon > 0$ and $\|\delta f\|_{\ell_1(\mathbb{N})} \leq \gamma$. Take $r \in \mathbb{N}$ such that

$$\mathbb{P}(\theta(\mathbf{U}) \geq r) < \epsilon/(8\gamma). \quad (14)$$

For $j \in \mathbb{Z}$, let $\mathbf{V}^{[j]}$ be a family of i.i.d. processes with values in \mathcal{U} independent of each other and of \mathbf{U} . We define a process \mathbf{Y} as

$$Y_j = [\Phi(V_{j+r}^{[j], \infty} U_{j-r}^{j+r} V_{-\infty}^{[j], j-r})]_j, \quad \text{for any } j \in \mathbb{Z}$$

where, for any i and j in \mathbb{Z} , we use the notation $V_i^{[j], \infty}$ for the sequence $V_i^{[j]}, V_{i+1}^{[j]}, \dots$. Clearly, \mathbf{Y} is stationary and if $\theta(\mathbf{U}) < r$, $Y_0 = X_0$. Moreover, \mathbf{Y} is a $2r$ -dependent process *i.e.*, for all $l, m > 1$ and $\underline{y} \in A^{\mathbb{Z}}$

$$\mathbb{P}(Y_1^l = y_1^l, Y_{l+2r+1}^{l+2r+m} = y_{l+2r+1}^{l+2r+m}) = \mathbb{P}(Y_1^l = y_1^l) \mathbb{P}(Y_{l+2r+1}^{l+2r+m} = y_{l+2r+1}^{l+2r+m}).$$

Now, observe that

$$f(X_1^n) - \mathbb{E}[f(X_1^n)] = f(X_j^n) - f(Y_1^n) - \mathbb{E}[f(X_1^n) - f(Y_1^n)] + f(Y_1^n) - \mathbb{E}[f(Y_1^n)]. \quad (15)$$

From the definition of f , we have that

$$f(X_1^n) - f(Y_1^n) \leq \sum_{j=1}^n \mathbf{1}\{X_j \neq Y_j\} \delta_j f \leq \sum_{j=1}^n \mathbf{1}\{\theta(T_{\mathcal{U}}^j \mathbf{U}) > r\} \delta_j f \quad (16)$$

and

$$|-\mathbb{E}[f(X_1^n) - f(Y_1^n)]| \leq \mathbb{P}(\theta(\mathbf{U}) > r) \sum_{j=1}^n \delta_j f \leq \epsilon/8. \quad (17)$$

Let $\alpha_j = \delta_j f / \|\delta f\|_{\ell_1(\mathbb{N})}$. Collecting (15), (16), (17), we have

$$\begin{aligned} & \mathbb{P}(|f(X_1^n) - \mathbb{E}[f(X_1^n)]| > \epsilon) \\ & \leq \mathbb{P} \left(\left| \sum_{j=1}^n \alpha_j \mathbf{1}\{\theta(T_{\mathcal{U}}^j \mathbf{U}) > r\} - \mathbb{P}(\theta(\mathbf{U}) > r) \right| > \frac{\epsilon}{4\|\delta f\|_{\ell_1(\mathbb{N})}} \right) + \mathbb{P}(|f(Y_1^n) - \mathbb{E}[f(Y_1^n)]| > \epsilon/2). \end{aligned} \quad (18)$$

We will use Lemma 2 to obtain upper bounds for the two terms of the right hand side of (18).

Let us begin with the second term, and we will use the fact that \mathbf{Y} is $2r$ -dependent process. First we define a coupling $(\tilde{\mathbf{Y}}, \hat{\mathbf{Y}})$, where $\tilde{\mathbf{Y}}$ and $\hat{\mathbf{Y}}$ are copies of \mathbf{Y} , by

$$(\tilde{Y}_j, \hat{Y}_j) := \left([\Phi(V_{j+r}^{[j], \infty} U_{j-r}^{j+r} V_{-\infty}^{[j], j-r})]_j, [\Phi(V_{j+r}'^{[j], \infty} U_{j-r}'^{j+r} V_{-\infty}'^{[j], j-r})]_j \right), \quad \text{for all } j \in \mathbb{Z}$$

where $\mathbf{V}^{[j]}, \mathbf{V}'^{[j]}$ and \mathbf{U}, \mathbf{U}' are i.i.d. processes satisfying the following properties. For $j > 0$, $\mathbf{V}^{[j]} = \mathbf{V}'^{[j]}$. For $j \leq 0$, $\mathbf{V}^{[j]}$ is independent of $\mathbf{V}'^{[j]}$, and both are independent of the rest. The processes \mathbf{U}, \mathbf{U}' are independent of $\mathbf{V}^{[j]}, \mathbf{V}'^{[j]}$ for all $j \in \mathbb{Z}$. Also $\{U_j\}_{j \leq r}$ and $\{U'_j\}_{j \leq r}$ are independent and for $j > r$, $U_j = U'_j$. From the construction, we have that, for $j > 2r$, $\mathbb{P}(\tilde{Y}_j = \hat{Y}_j) = 1$. Moreover, for all $\underline{y}, \underline{z} \in A^{-\mathbb{N}}$ and $j > 2r$,

$$\mathbb{P}(\tilde{Y}_j = \hat{Y}_j | \tilde{Y}_{-\infty}^0 = \underline{y}, \hat{Y}_{-\infty}^0 = \underline{z}) = \mathbb{P}(\tilde{Y}_j = \hat{Y}_j).$$

Therefore, using the above coupling and Lemma 2, we have

$$\mathbb{P}(|f(Y_1^n) - \mathbb{E}[f(Y_1^n)]| > \epsilon/2) \leq 2 \exp\left(-\frac{g_1(\epsilon, \gamma)}{\|\delta f\|_{\ell_2(\mathbb{N})}^2}\right)$$

where $g_1(\epsilon, \gamma) = \frac{1}{2}\epsilon^2(1 + r + \sum_{j=1}^r \mathbb{P}(\theta(\mathbf{U}) \geq j))^{-2}$. Note that r depends on ϵ and γ .

For the first term of the right hand side of (18), let $Z_j = \mathbf{1}\{\theta(T_{\mathcal{U}}^j \mathbf{U}) \geq r\}$. Observe that the process $\mathbf{Z} = \{Z_j\}_{j \in \mathbb{Z}}$ is stationary. Also because the event $\{\theta(\mathbf{U}) < r\}$ is $\mathcal{F}(U_{-r}^r)$ measurable, the process \mathbf{Z} is a $2r$ -dependent process. We define a coupling $(\tilde{\mathbf{Z}}, \hat{\mathbf{Z}})$, where $\tilde{\mathbf{Z}}$ and $\hat{\mathbf{Z}}$ are copies of \mathbf{Z} , by

$$(\tilde{Z}_j, \hat{Z}_j) = (\mathbf{1}\{\theta(T_{\mathcal{U}}^j \mathbf{U}) \geq r\}, \mathbf{1}\{\theta(T_{\mathcal{U}'}^j \mathbf{U}') \geq r\})$$

where \mathbf{U} and \mathbf{U}' are i.i.d. process with $U_j = U'_j$ for $j > r$ and $\{U_j\}_{j \leq r}$ and $\{U'_j\}_{j \leq r}$ are independent. Therefore, again by Lemma 2 we have that

$$\mathbb{P}\left(\left|\sum_{j=1}^n \alpha_j \mathbf{1}\{\theta(T_{\mathcal{U}}^j \mathbf{U}) > r\} - \mathbb{P}(\theta(\mathbf{U}) > r)\right| > \frac{\epsilon}{4\|\delta f\|_{\ell_1(\mathbb{N})}}\right) \leq 2 \exp\left(-\frac{g_2(\epsilon, \gamma)}{\|\delta f\|_{\ell_2(\mathbb{N})}^2}\right), \quad (19)$$

where $g_2(\epsilon, \gamma) = \frac{1}{8}\epsilon^2(1 + r + \sum_{j=1}^r \mathbb{P}(\theta(\mathbf{U}) \geq j))^{-2}$. Thus we have

$$\mathbb{P}(|f(X_1^n) - \mathbb{E}[f(X_1^n)]| > \epsilon) \leq 4 \exp\left(-\frac{g_2(\epsilon, \gamma)}{\|\delta f\|_{\ell_2(\mathbb{N})}^2}\right), \quad (20)$$

which proves the theorem.

6.3. Proof of Proposition 1. We use the same notation as in the proof of Theorem 3. We define a coupling $(\tilde{\mathbf{X}}, \hat{\mathbf{X}})$ between two copies of \mathbf{X} by

$$(\tilde{X}_j, \hat{X}_j) := ([\Phi(\mathbf{U})]_j, [\Phi(\mathbf{U}')]_j), \text{ for all } j \in \mathbb{Z}$$

where \mathbf{U} and \mathbf{U}' are i.i.d. process such that, for $j > 0$, $U_j = U'_j$ and $\{U_j\}_{j \leq 0}$ and $\{U'_j\}_{j \leq 0}$ are independent. From the definition of CFTP algorithm, we can take $\Delta \leq \sum_{j=1}^{\infty} \mathbb{P}(\theta(\mathbf{U}) > j) = \mathbb{E}[\theta]$. Applying Lemma 2, we obtain (5).

6.4. Proof of Theorem 4. We say that a stationary process \mathbf{X} has the *positive divergence property* if

$$\liminf_n \frac{1}{n+1} \sum_{a \in \{-1, 1\}^{n+1}} \mathbb{P}(Y_{-n}^0 = a_{-n}^0) \log \frac{\mathbb{P}(Y_{-n}^0 = a_{-n}^0)}{\mathbb{P}(X_{-n}^0 = a_{-n}^0)} > 0$$

for any ergodic process \mathbf{Y} different of \mathbf{X} .

The proof of Theorem 4 is based on the following lemmas.

Lemma 3 (Marton & Shields (1994)). *Let \mathbf{X} be a stationary process and let the ergodic theorem holds at exponential rate. Then \mathbf{X} has the positive divergence property.*

Proof. If \mathbf{X} has the concentration of measure at exponential rate it satisfies (4). Therefore, Theorem 1 of Marton & Shields (1994) implies that \mathbf{X} has the positive divergence property. \square

Lemma 4. *Let \mathbf{X} and \mathbf{Y} be two stationary processes compatible with a continuous kernel P . Let $\inf_{a \in \{-1,1\}^{\mathbb{Z}}} P(a_0 | a_{-\infty}^{-1}) = \delta > 0$. Then the relative entropy rate*

$$\lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{a \in \{-1,1\}^{n+1}} \mathbb{P}(X_{-n}^0 = a_{-n}^0) \log \frac{\mathbb{P}(X_{-n}^0 = a_{-n}^0)}{\mathbb{P}(Y_{-n}^0 = a_{-n}^0)}$$

exists and is 0.

Proof. Let \mathbf{Z} represent \mathbf{X} or \mathbf{Y} . Define, for $k \in \mathbb{N}$,

$$H_{\mathbf{X}}(Z_{-k}^0) = - \sum_{a \in \{-1,1\}^{k+1}} \mathbb{P}(X_{-k}^0 = a_{-k}^0) \log \mathbb{P}(Z_{-k}^0 = a_{-k}^0).$$

Now, we can rewrite the relative entropy rate as

$$\lim_{n \rightarrow \infty} \frac{1}{n+1} \{H_{\mathbf{X}}(Y_{-n}^0) - H_{\mathbf{X}}(X_{-n}^0)\}.$$

Define also, for $k \in \mathbb{N}^*$,

$$H_{\mathbf{X}}(Z_0 | Z_{-k}^{-1}) = - \sum_{a \in \{-1,1\}^{k+1}} \mathbb{P}(X_{-k}^0 = a_{-k}^0) \log \mathbb{P}(Z_0 = a_0 | Z_{-k}^{-1} = a_{-k}^{-1}).$$

By the chain rule and stationarity of the processes, we have

$$H_{\mathbf{X}}(Z_{-k}^0) = H_{\mathbf{X}}(Z_0) + \sum_{k=1}^n H_{\mathbf{X}}(Z_0 | Z_{-k}^{-1}).$$

Therefore, we have that the relative entropy rate is a difference between two Cesàro sums. To prove that the relative entropy exists, it is enough to show that the limit

$$\lim_{n \rightarrow \infty} H_{\mathbf{X}}(Z_0 | Z_{-n}^{-1})$$

exists. To see that $H_{\mathbf{X}}(Z_0 | Z_{-n}^{-1})$ converges, let $\mu_{\mathbf{Z}}$ be the measure associated with \mathbf{Z} , we have

$$H_{\mathbf{X}}(Z_0 | Z_{-n}^{-1}) = -\mathbb{E}_{\mathbf{X}}(\log(\mu_{\mathbf{Z}}(X_0 | X_{-n}^{-1}))).$$

By assumption, for all $a \in \{-1,1\}^{\mathbb{Z}}$

$$-\log(\mu_{\mathbf{Z}}(a_0 | a_{-n}^{-1})) \leq -\log \delta,$$

therefore, by dominated convergence theorem

$$\lim_{n \rightarrow \infty} -\mathbb{E}_{\mathbf{X}}(\log(\mu_{\mathbf{Z}}(X_0 | X_{-n}^{-1}))) = -\mathbb{E}_{\mathbf{X}}(\lim_{n \rightarrow \infty} \log(\mu_{\mathbf{Z}}(X_0 | X_{-n}^{-1}))).$$

By continuity of P we have that, for all $a \in \{-1,1\}^{\mathbb{Z}}$

$$\lim_{n \rightarrow \infty} \log(\mu_{\mathbf{Z}}(a_0 | a_{-n}^{-1})) = \log P(Y_0 = a_0 | Y_{-\infty}^{-1} = a_{-\infty}^{-1}).$$

Hence,

$$\lim_{n \rightarrow \infty} H_{\mathbf{X}}(Z_0 | Z_{-n}^{-1}) = -\mathbb{E}_{\mathbf{X}}(\log(P(X_0 | X_{-\infty}^{-1}))),$$

which concludes the proof. \square

Proof of Theorem 4. If \mathbf{X} has the concentration of measure holding at exponential rate, by Lemma 3, \mathbf{X} has the positive divergence property. By Lemma 4, if \mathbf{X} has the positive divergence property, there is no other ergodic process compatible with P . This concludes the proof. \square

6.5. **Proof of Theorem 5.** The kernel of the original BK example (see Section 2) writes as follows

$$P^{[m_k]}(+1|a_{-m_k}^{-1}) := \left((1-\epsilon)\mathbf{1}\left\{\frac{1}{m_k}\sum_{i=1}^{m_k} a_{-i} > 0\right\} + \epsilon\mathbf{1}\left\{\frac{1}{m_k}\sum_{i=1}^{m_k} a_{-i} < 0\right\} \right), \quad (21)$$

with $\lambda_k = cr^k$ for some positive constant and $r \in (2/3, 1)$. In other words, the BK example is already given under the form of a countable mixture of Markov kernels $P^{[m_k]}$, $k \geq 1$ of lacunary ranges. Here, the parameter of the kernel is the sequence $\{m_k\}_{k \geq 1}$, and taking it increasing sufficiently fast, Bramson & Kalikow (1993) proved that there is phase transition. On the other hand, taking this sequence increasing slowly, the resulting kernel will satisfy uniqueness and we denote by \mathbf{X}^{BK} the unique stationary chain compatible with P^{BK} in this case.

We will prove Theorem 5 in two steps. First, Lemma 5 states that \mathbf{X}^{BK} is a finitary coding of an i.i.d. process with discrete state space and finite entropy. Then, Lemma 6 (stated and proved in Appendix A) states that i.i.d. chains with finite entropy are finitary codings of i.i.d. chains on a sufficiently large, yet finite, alphabet. This later result is known (Rudolph, 1982), but our proof is new and simple, and we think that it is worth mentioning here. Since it has a life of its own, we decided to put the statement and proof of this lemma in an appendix.

We will conclude the proof of the theorem observing that if a process \mathbf{X} is a finitary coding of a process \mathbf{Y} , which is itself a finitary coding of a process \mathbf{Z} , then \mathbf{X} is also a finitary coding of \mathbf{Z} .

Let \mathbf{L}^{BK} be an i.i.d. process taking value in the alphabet $A := \{0_-, 0_+\} \cup \{1, 2, \dots\}$, with distribution $Q(L_0^{BK} = 0_-) = Q(L_0^{BK} = 0_+) = \epsilon$ and $Q(L_0^{BK} = k) = \lambda_k(1 - 2\epsilon)$ for $k \geq 1$. Observe that \mathbf{L}^{BK} has finite entropy, since

$$H(\mathbf{L}^{BK}) := \sum_{a \in A} Q(a) \log Q(a) = 2\epsilon \log \epsilon + (1 - 2\epsilon) \log(1 - 2\epsilon) + (1 - 2\epsilon) \sum_{k \geq 1} \lambda_k \log \lambda_k$$

is finite for the sequence $\{\lambda_k\}_{k \geq 1}$ we consider here (in the original BK example).

Lemma 5. \mathbf{X}^{BK} is a finitary coding of \mathbf{L}^{BK} .

Proof. We first rewrite the kernel P^{BK} in the following way

$$P^{BK}(+1|\underline{a}) = \epsilon + \sum_{k \geq 1} (1 - 2\epsilon) \lambda_k \mathbf{1}\left\{\frac{1}{m_j} \sum_{i=1}^{m_j} a_{-j} > 0\right\}. \quad (22)$$

This decomposition means that the construction of the next symbol of the chain using P^{BK} can be done using the following two steps procedure: (1) generate L_0^{BK} with distribution Q and (2) if $L_0^{BK} = 0_-$ or 0_+ , put -1 or $+1$, otherwise, if $L_0^{BK} = k \geq 1$, put the symbol that most occurs in $X_{-m_k}^{-1}$. This procedure motivates the use of the following update function for the coupling between the chain with fixed past $-\underline{1}$ and the chain with fixed past $+\underline{1}$:

$$F^{BK}((\underline{a}, \underline{b}), L_0^{BK}) := \begin{cases} (-1, -1) & \text{if } L_0^{BK} = 0_- \\ (+1, +1) & \text{if } L_0^{BK} = 0_+ \\ (\text{maj}(a_{-m_k}^{-1}), \text{maj}(b_{-m_k}^{-1})) & \text{if } L_0^{BK} = k \geq 1 \end{cases} \quad (23)$$

where $\text{maj}(a_{-m_k}^{-1})$ denotes the symbol of $\{-1, +1\}$ that most occurs in $a_{-m_k}^{-1}$. This update function defines a coupling kernel which satisfies (6) for any pairs of ordered pasts $\underline{a} \leq \underline{b}$, and therefore, can be used to perform the Attractive Sampler. It follows, using the same proof as the one of Theorem 2 that \mathbf{X}^{BK} is a finitary coding from \mathbf{L}^{BK} via the Attractive Sampler. \square

APPENDIX A. FINITARY CODING FROM A FINITE ALPHABET I.I.D. PROCESS TO A COUNTABLE ALPHABET I.I.D. PROCESS

Lemma 6. *An \mathbb{N} -valued i.i.d. process having finite entropy can be constructed as a finitary coding of an i.i.d. process with finite alphabet.*

We will explicitly construct the coding from a sequence \mathcal{P} of *piles* of (a sufficiently large integer we will fix) N (unbiased random binary) bits, that is, we construct a finitary coding from an i.i.d. process taking values (uniformly) in $\{0, 1\}^N$ to the \mathbb{N} -valued i.i.d. process we will denote \mathbf{Y} . This algorithm is in the spirit of the one of Harvey *et al.* (2007). The proof that this coding is actually finitary uses the techniques developed in Ferrari *et al.* (2000). In fact, we will show more: the coding length has finite expectation.

Before we present the algorithm, we need some more definitions.

A *simulation* of $\{q(k)\}_{k \geq 0} = \{\nu_k\}_{k \geq 0}$ (q distribution on \mathbb{N}) from independent unbiased bits (law \mathbb{P}) is a pair of measurable functions $T : \{0, 1\}^{\mathbb{N}} \rightarrow \mathbb{N}$ and $S : \{0, 1\}^{\mathbb{N}} \rightarrow \mathbb{N}$ defined \mathbb{P} -a.s., with the following properties:

- (1) If $a_1^{+\infty}, \tilde{a}_1^{+\infty} \in \{0, 1\}^{\mathbb{N}}$ are such that $\tilde{a}_1 \dots \tilde{a}_{T(a_1^{+\infty})} = a_1 \dots a_{T(a_1^{+\infty})}$, then $T(a_1^{+\infty}) = T(\tilde{a}_1^{+\infty})$ and $S(a_1^{+\infty}) = S(\tilde{a}_1^{+\infty})$.
- (2) Under the measure \mathbb{P} , $S(a_1^{+\infty})$ has distribution q .

T is called the stopping time of the simulation, and S is called the output symbol. We do not give the explicit definition of such a pair of functions in order to save space. We refer the interested reader to the papers of Knuth & Yao (1976), Romik (1999) and Harvey *et al.* (2007) for further details. For any given distribution q , there is a fixed $m^* \geq 1$ such that $T(a_1^{+\infty}) \geq m^*$ for any $a_1^{+\infty}$. On the other hand, we have the following result which follows, for example, from Item (ii) of Theorem 1 in Romik (1999).

Lemma 7. *Assume that $\sum_{k \geq 0} \nu_k \log \nu_k < +\infty$, then, there exists pair (T, S) such that*

$$\mathbb{E}T := \mathbb{E}T(a_1^{+\infty}) < +\infty.$$

For the piles at each sites, we use the notation $\mathcal{P}_k := a_1^{(k)}, \dots, a_N^{(k)}$, where $a_i^{(k)} \in \{0, 1\}$ for any $i \in \{1, \dots, N\}$ and $k \in \mathbb{Z}$. For any $k \in \mathbb{Z}$, let us denote by (T_k, S_k) the simulation of site k . The sequence of bits read by (T_k, S_k) is a subset of the piles of bits \mathcal{P}_i , $i \leq k$. This is what we explain now, first in an informal way (borrowed from Harvey *et al.* (2007)).

Observation 2. *When we say that we use the bits of a pile, we mean we use them from the bottom to the top. If we use bits of a pile which has already been used, we begin from the lower unused one, and climb up the pile.*

Algorithm inspired on Harvey *et al.* (2007). For each site $k \in \mathbb{Z}$, (T_k, S_k) attempts to use the random bits in \mathcal{P}_k to simulate Y_k . In many sites $k \in \mathbb{Z}$, the stopping time T_k will be reached, that is $T_k(a_1^{(k)} \dots a_N^{(k)} a_{N+1}^{+\infty}) \leq N$ for any $a_{N+1}^{+\infty}$. If the stopping time is reached, \mathcal{P}_k may contain unused bits, which are still independent and unbiased, that is, if $T_k(a_1^{(k)} \dots a_N^{(k)} a_{N+1}^{+\infty}) = i \leq N$, then $a_{i+1}^{(k)} \dots a_N^{(k)}$ are still unused, independent and unbiased. For any site k whose stopping time T_k is not reached, look at \mathcal{P}_{k-1} in the *next site to the left* (that is, the simulation goes backwards in time) to find unused bits to continue the simulation. If now the stopping time T_k is reached, compute Y_k . If not, iterate, looking one site further to the left at each step for unused bits, until the

stopping time is reached. This iteration is done simultaneously for all sites, in order to maintain translation-equivariance of the construction.

More formally, the string of bits read by (T_k, S_k) , $k \geq 0$, when we limit ourselves to looking at the piles \mathcal{P}_i , $i \geq 0$ are defined as follow:

$$\begin{aligned} z^{0 \rightarrow 0} &:= a_1^{(0)} \dots a_N^{(0)} = z_1^{0 \rightarrow 0} \dots z_N^{0 \rightarrow 0} \\ z^{0 \rightarrow 1} &:= a_1^{(1)} \dots a_N^{(1)} z_{T_0(z^{0 \rightarrow 0})+1}^{0 \rightarrow 0} \dots z_{|z^{0 \rightarrow 0}|}^{0 \rightarrow 0} \\ z^{0 \rightarrow 2} &:= a_1^{(2)} \dots a_N^{(2)} z_{T_1(z^{0 \rightarrow 1})+1}^{0 \rightarrow 1} \dots z_{|z^{0 \rightarrow 1}|}^{0 \rightarrow 1} \\ &\dots \\ z^{0 \rightarrow k} &:= a_1^{(k)} \dots a_N^{(k)} z_{T_{k-1}(z^{0 \rightarrow k-1})+1}^{0 \rightarrow k-1} \dots z_{|z^{0 \rightarrow k-1}|}^{0 \rightarrow k-1} \end{aligned}$$

where, for any $k \geq 0$, the length of the string $z^{0 \rightarrow k}$ is denoted by $|z^{0 \rightarrow k}|$, and we used the convention that $T_k(z^{0 \rightarrow k}) = \infty$ when the simulation is not successful using only the string $z^{0 \rightarrow k}$, otherwise, $T_k(z^{0 \rightarrow k}) = i \leq |z^{0 \rightarrow k}|$, the smallest integer such that $T_k(z^{0 \rightarrow k} b_1^{+\infty}) = i$ for any $b_1^{+\infty}$.

We now define

$$\mathcal{T}(\mathcal{P})[0] := \max\{j \leq 0 : T_i(z^{j \rightarrow i}) \leq |z^{j \rightarrow i}|, i = j, \dots, 0\}, \quad (24)$$

and

$$[S(\mathcal{P})]_0 := S(z^{\mathcal{T}[0] \rightarrow 0}).$$

When $\mathcal{T}(\mathcal{P})[0]$ is a.s. finite, the construction is feasible, and the algorithm constructs a sample of the target distribution, that is the product measure $q^{\mathbb{Z}}$. This is because the strings of unbiased bits $z^{\mathcal{T}[i] \rightarrow i}$, $i \in \mathbb{Z}$ are all finite and disjoint, and the simulators (T_i, S_i) , $i \in \mathbb{Z}$ are designed to read strings of unbiased bits and transform them into bits distributed according to q . Thus, in order to prove Lemma 6, it is enough to prove that $\mathcal{T}(\mathcal{P})[0]$ is a.s. finite.

Proof of Lemma 6. Let us first show that

$$\mathbb{P}(\mathcal{E}) := \mathbb{P}(T_0(z^{0 \rightarrow 0}) \leq |z^{0 \rightarrow 0}|, T_1(z^{0 \rightarrow 1}) \leq |z^{0 \rightarrow 1}|, \dots, T_j(z^{0 \rightarrow j}) \leq |z^{0 \rightarrow j}|, \dots) > 0.$$

Observe that on the event \mathcal{E} , all the strings $z^{0 \rightarrow i}$ are finite. More specifically, we have $|z^{0 \rightarrow 0}| \leq N$ and $|z^{0 \rightarrow j}| \leq (j+1)N - \sum_{i=0}^{j-1} T_i(z^{0 \rightarrow i})$, for any $j \geq 1$. Thus,

$$\mathbb{P}(\mathcal{E}) = \mathbb{P}\left(T_0(z^{0 \rightarrow 0}) \leq N, T_0(z^{0 \rightarrow 0}) + T_1(z^{0 \rightarrow 1}) \leq 2N, \dots, \sum_{i=0}^j T_i(z^{0 \rightarrow i}) \leq (j+1)N, \dots\right)$$

But $\{z^{0 \rightarrow j}\}_{j \geq 0}$ constitutes a sequence of disjoint strings of unbiased independent random bits. For this reason, for any sequence of independent infinite strings of binary bits $\{a^{(i)}\}_{i \geq 0}$, $a^{(i)} := a_0^{(i)} a_1^{(i)} \dots$, we have

$$\mathbb{P}(\mathcal{E}) = \mathbb{P}\left(T_0(z^{0 \rightarrow 0} a^0) \leq N, T_0(z^{0 \rightarrow 0} a^0) + T_1(z^{0 \rightarrow 1} a^1) \leq 2N, \dots, \sum_{i=0}^j T_i(z^{0 \rightarrow i} a^i) \leq (j+1)N, \dots\right)$$

Now, $\{T_j(z^{0 \rightarrow j} a^j)\}_{j \geq 0}$ is a sequence of i.i.d. random variables taking values in $\{m^*, m^*+1, \dots\}$ (we recall that m^* is the smallest number of bits necessary to simulate from q), and having expectation $\mathbb{E}T_0(z^{0 \rightarrow 0} a^0) < N$. Denoting $\zeta_j := T_j(z^{0 \rightarrow j} a^j) - N$, we obtain an i.i.d. sequence ζ (where $\zeta_i \in \{-N + m^*, \dots\}$) such that $\mathbb{E}\zeta < 0$. Therefore

$$\mathbb{P}(\mathcal{E}) = \mathbb{P}(\zeta_1 \leq 0, \zeta_1 + \zeta_2 \leq 0, \zeta_1 + \zeta_2 + \zeta_3 \leq 0, \dots) > 0$$

is true because $\sum_{i=1}^k \zeta_i$ is a *negatively drifted random walk*, which is transient.

Now, define the chain ξ on $\{0, 1\}$ by $\xi_j := \mathbb{1}\{j = \mathcal{T}[j, +\infty]\}$. Then, consider the sequence of time indexes \mathbf{T} defined by $\xi_j = 1$ if and only if $j = T_l$ for some l in \mathbb{Z} , $T_l < T_{l+1}$ and with the convention $T_0 \leq 0 < T_1$. We will now prove that the chain ξ is renewal (that is, the increments $\{T_{i+1} - T_i\}_{i \in \mathbb{Z}}$ are independent, and are identically distributed for $i \neq 0$ with finite expectation). Define for any $-\infty < m \leq n \leq +\infty$ the events

$$H[m, n] := \{T_i(z^{m \rightarrow i}) \leq |z^{m \rightarrow i}|, i \in \{m, \dots, n\}\}.$$

which is measurable with respect to the σ -algebra generated by \mathcal{P}_m^n . By definition (24) of $\mathcal{T}[m, n]$, $-\infty < m \leq n \leq +\infty$, we have for any $t_1 < t_2 < \dots < t_n$

$$\bigcap_{l=1}^n \{\mathcal{T}[t_l, +\infty] = t_l\} = \bigcap_{l=1}^n H[t_l, t_{l+1} - 1] \quad (25)$$

where $t_{n+1} := +\infty$. This is an intersection of independent events. Then, we observe that by stationarity,

$$\mathbb{P}(H[j, +\infty]) = \mathbb{P}(H[0, +\infty]) = \mathbb{P}(\mathcal{E}) > 0$$

and

$$\mathbb{P}(H[-j, -1]) = \mathbb{P}(H[-j, +\infty] | H[0, +\infty]), \quad \forall j \geq 1.$$

Together with (25), this yields for any sequence of integers $t_1 < t_2 < \dots < t_n$

$$\mathbb{P}(\xi_{t_l} = 1, l = 1, \dots, n) = \mathbb{P}(\mathcal{E}) \prod_{l=1}^{n-1} \mathbb{P}(\xi_{-(t_{l+1})-t_l} = 1 | \xi_0 = 1)$$

and therefore, the chain ξ is renewal (has positive renewal probability). It follows that $\{T_{i+1} - T_i\}_{i \in \mathbb{Z}}$ are independent, and are identically distributed for $i \neq 0$ with finite expectation $\mathbb{P}(\mathcal{E})^{-1}$. We conclude that $\mathcal{T}[0]$ has finite expectation since

$$\mathbb{P}(T_{l+1} - T_l \geq m) = \mathbb{P}(\mathcal{T}[1, +\infty] \leq -m | \mathcal{T}[0, +\infty] = 0) = \mathbb{P}(\mathcal{T}[0] < -m + 1).$$

□

Acknowledgments. We gratefully acknowledge A. Galves, R. Fernández, K. Marton and S. Friedli for many discussion during the elaboration of this paper.

REFERENCES

- BERGER, N., HOFFMAN, C. & SIDORAVICIUS, V. (2005). Nonuniqueness for specifications in $\ell^{2+\epsilon}$. *arXiv:math/0312344*.
- BRAMSON, M. & KALIKOW, S. (1993). Nonuniqueness in g -functions. *Israel J. Math.* **84**(1-2), 153–160.
- CÉNAC, P., CHAUVIN, B., PACCAUT, F. & POUYANNE, N. (2010). Variable length markov chains and dynamical sources. *HAL : hal-00503150*.
- CHAZOTTES, J., COLLET, P., KUELSKE, C. & REDIG, F. (2007). Concentration inequalities for random fields via coupling. *Probability Theory and Related Fields* **137**, 201225.
- COMETS, F., FERNANDEZ, R. & FERRARI, P. (2002). Processes with long memory: regenerative construction and perfect simulation. *Ann. Appl. Prob.* **12**(3), 921–943.
- DE SANTIS, E. & PICCIONI, M. (2010). A general framework for perfect simulation of long memory processes. *arXiv:1004.0113v1*.

- DOEBLIN, W. & FORTET, R. (1937). Sur des chaînes à liaisons complètes. *Bull. Soc. Math. France* **65**, 132–148.
- FERNÁNDEZ, R., GALLO, S. & MAILLARD, G. (2011). Regular g -measure are not always gibbsian. *Accepted for publication in Elec. Comm. Probab.* .
- FERNÁNDEZ, R. & MAILLARD, G. (2005). Chains with complete connections: general theory, uniqueness, loss of memory and mixing properties. *J. Stat. Phys.* **118**(3-4), 555–588. URL <http://dx.doi.org/10.1007/s10955-004-8821-5>.
- FERRARI, P. A., MAASS, A., MARTÍNEZ, S. & NEY, P. (2000). Cesàro mean distribution of group automata starting from measures with summable decay. *Ergodic Theory Dynam. Systems* **20**(6), 1657–1670.
- FRIEDLI, S. (2010). A note on the bramson-kalikow process .
- GALLO, S. & GARCIA, N. L. (2011). General context-tree-based approach to perfect simulation for chains of infinite order. *Submitted, arXiv:1103.2058v2* .
- HARRIS, T. E. (1955). On chains of infinite order. *Pacific J. Math.* **5**, 707–724.
- HARVEY, N., HOLROYD, A. E., PERES, Y. & ROMIK, D. (2007). Universal finitary codes with exponential tails. *Proc. Lond. Math. Soc. (3)* **94**(2), 475–496. URL <http://dx.doi.org/10.1112/plms/pdl018>.
- HULSE, P. (1991). Uniqueness and ergodic properties of attractive g -measures. *Ergodic Theory and Dynamical Systems* **11**, 65–77.
- HULSE, P. (2006). An example of non-unique g -measures. *Ergodic Theory Dynam. Systems* **26**(2), 439–445. URL <http://dx.doi.org/10.1017/S0143385705000489>.
- JOHANSSON, A. & ÖBERG, A. (2003). Square summability of variations of g -functions and uniqueness of g -measures. *Math. Res. Lett.* **10**(5-6), 587–601.
- KALIKOW, S. (1990). Random Markov processes and uniform martingales. *Israel J. Math.* **71**(1), 33–54. URL <http://dx.doi.org/10.1007/BF02807249>.
- KARLIN, S. (1953). Some random walks arising in learning models i. *Pacific Journal of Mathematics* **3**, 725–756.
- KEANE, M. (1972). Strongly mixing g -measures. *Invent. Math.* **16**, 309–324.
- KNUTH, D. E. & YAO, A. C. (1976). The complexity of nonuniform random number generation. In: *Algorithms and complexity (Proc. Sympos., Carnegie-Mellon Univ., Pittsburgh, Pa., 1976)*. New York: Academic Press, pp. 357–428.
- LACROIX, Y. (2000). A note on weak- \star perturbations of g -measures. *Sankhyā Ser. A* **62**(3), 331–338. Ergodic theory and harmonic analysis (Mumbai, 1999).
- MARTON, K. & SHIELDS, P. (1994). The positive divergence and blowing up properties. *Israeli Journal of Mathematics* **86**, 331–348.
- MCCULLAGH, P. & NELDER, J. A. (1983). *Generalized linear models*. Monographs on Statistics and Applied Probability. London: Chapman & Hall.
- ONICESCU, O. & MIHOC, G. (1935). Sur les chaînes de variables statistiques. *Bull. Sci. Math* **59**(2), 174–192.
- PETIT, B. (1982). Deux schémas de bernoulli d’alphabet dénombrable et d’entropie infinie sont finitairement isomorphes. *Z. Wahrsch. Verw. Gebiete* **59**, 161–168.
- PRESTON, C. (1976). *Random Fields*. Lecture Notes in Mathematics (Springer-Verlag). Berlin: Springer Verlag.

- PROPP, J. G. & WILSON, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. In: *Proceedings of the Seventh International Conference on Random Structures and Algorithms (Atlanta, GA, 1995)*, vol. 9.
- QUAS, A. (1996). Non-ergodicity for c_1 expanding maps and g -measures. *Ergodic Theory and Dynamical Systems* **16**, 531–543.
- ROMIK, D. (1999). Sharp entropy bounds for discrete statistical simulation. *Statistics and Probability Letters* **42**, 219227.
- RUDOLPH, D. J. (1982). A mixing Markov chain with exponentially decaying return times is finitarily Bernoulli. *Ergodic Theory Dynamical Systems* **2**(1), 85–97.
- SHIELDS, P. C. (1996). *The ergodic theory of discrete sample paths*, vol. 13 of *Graduate Studies in Mathematics*. Providence, RI: American Mathematical Society.
- STEIF, J. & VAN DEN BERG, J. (1999). On the existence and nonexistence of finitary codings for a class of random fields. *Annals of Probability* **11**, 1501–1522.
- STENFLO, O. (2001). A note on a theorem of karlin. *Statistics and Probability Letters* **54**, 183187.

INSTITUTO DE MATEMÁTICA, UNIVERSIDADE FEDERAL DE RIO DE JANEIRO
E-mail address: `sandro@im.ufrj.br`

PRINCETON UNIVERSITY, GREEN HALL, NJ, 08540
E-mail address: `takahashi@d@gmail.com`